# CLAIMS

What is claimed is:

1. An information retrieval and document navigation system, comprising:

a datastore of direct links between pre-defined core concepts found in a document corpus;

a link identification module adapted to identify indirect links between core concepts selected by a user based on connection of direct links through at least one core concept not selected by the user; and

an output adapted to communicate identified links to the user.

2. The system of claim 1, further comprising a co-occurrence detection module finding the direct links by detecting co-occurrence between core concepts in the document corpus and employing a mutual information technique including the Fisher exact test to obtain a statistical P value expressing a significance of a detected co-occurrence.

3. The system of claim 2, wherein said co-occurrence detection module is adapted to identify an alias of a core concept in document contents, and to equate occurrence of the alias with occurrence of the core concept.

4. The system of claim 1, wherein said datastore further maintains pointers between detected co-occurrences and documents in which the co-occurrences are detected.

5. The system of claim 1, wherein said output is adapted to provide pointers to documents to the user, wherein the documents relate to an identified link.

6. The system of claim 1, further comprising multiple, discipline-focused lexica organized according to the core concepts and identifying aliases by which the core concepts may be found in document contents.

7. The system of claim 1, further comprising a user interface adapted to communicate selectable lexica to the user, to receive lexicon selections and initial search terms from the user, to extract aliases from the initial search terms, to identify candidate core concepts in lexica selected by the user based on the extracted aliases, and to communicate the candidate core concepts to the user for final selection.

8. The system of claim 1, further comprising an input receiving core concept selections and a specified depth of link from a user.

9. The system of claim 1, wherein said datastore is adapted to record a type of a link between two core concepts, wherein the type of link is automatically identified based on automatic detection in link-related document contents of one of plural, predefined, candidate relationships between predefined categories associated with the two core concepts.

10. The system of claim 1, wherein said datastore is adapted to record a direction of a link between two core concepts, wherein the direction of the link is automatically determined based on a type of the link between the two core concepts and predefined categories associated with the two core concepts.

11. The system of claim 1, wherein said output is adapted to communicate identified links to the user in the form of a matrix relating core concepts to core concepts.

12. The system of claim 1, further comprising a browsable lexicon of core concepts permitting the user to browse core concepts according to relationships between the core concepts and to select core concepts.

13. The system of claim 1, further comprising a pre-computed link datastore containing directional links between core concepts forming an extendable, searchable concept map in addition to manually curated links and supporting documents.

14. The system of claim 1, further comprising a datastore of curated relationships and automatically detected relationships between core concepts, wherein said output is adapted to at least one of:

(a) identify curated relationships as curated; and

(b) identify only curated relationships associated with a core concept based on user preference.

15. The system of claim 1, a plurality of links between biological sequence data and related documents in the document corpus.

16. An information retrieval and document navigation system, comprising:

multiple, discipline-focused lexica organized according to core concepts and identifying aliases by which the core concepts may be found in document contents;

a datastore of direct links between pre-defined core concepts found in a document corpus, wherein said datastore further maintains pointers between detected co-occurrences and documents in which the co-occurrences are detected;

a co-occurrence detection module finding the direct links by detecting co-occurrence between core concepts in the document corpus by employing a mutual information technique to obtain a level of statistical significance of a detected co-occurrence, wherein said co-occurrence detection module is adapted to identify an alias of a core concept in document contents, and to equate occurrence of the alias with occurrence of the core concept;

a link identification module adapted to identify indirect links between core concepts selected by a user based on connection of direct links through at least one core concept not selected by the user; and

an output adapted to communicate identified links and related pointers to documents supporting the identified links to the user.

17. The system of claim 16, wherein said output is adapted to render a graphic display of links between core concepts, with nodes corresponding to core concepts and edges corresponding to links.

18. The system of claim 17, wherein the nodes serve as hyperlinks to summaries of information relating to associated core concepts.

19. The system of claim 17, wherein the edges serve as hyperlinks to collections of pointers to documents supporting associated links.

20. The system of claim 17, wherein the edges have visual characteristics identifying at least one of a strength of relationship between bounding nodes, a type of relationship between bounding nodes, and a direction of relationship between bounding nodes.

21. The system of claim 16, further comprising a link relation module adapted to select a constraint list of candidate relationship types based on predefined categories associated with two core concepts bounding a direct link, and to automatically identify a type of relationship associated with the direct link by finding occurrences of constraint list elements in proximity to detected co-occurrences of the two core concepts in document contents supporting the direct link.

22. The system of claim 21, wherein the two core concepts are of different predefined categories, the candidate relationship types have a predefined direction between the two core concepts, and said link relation module is adapted to apply the predefined direction of the type of relationship associated with the direct link to the direct link.

23. The system of claim 21, wherein the two core concepts are of identical predefined categories, the candidate relationship types have predefined semantic templates adapted to identify directions between the two core concepts in document contents supporting the direct link, and said link relation module is adapted to automatically identify a direction associated with the direct link by matching a template of the type of relationship associated with the direct link to document contents in proximity to detected co-occurrences of the two core concepts in document contents supporting the direct link.

24. The system of claim 16, wherein said multiple, discipline-focused lexica include a gene lexicon organized according to core concepts corresponding to at least one of gene functions, protein functions, gene names, protein names, gene structures, and protein structures.

25. The system of claim 24, wherein said multiple, discipline-focused lexica include a disease lexicon, a drug lexicon, a tissue lexicon, and a taxonomy lexicon.

26. The system of claim 16, wherein the mutual information technique includes the Fisher exact test.

27. A method of information retrieval and document navigation, comprising:

finding direct links between pre-defined core concepts in a document corpus;

identifying indirect links between core concepts selected by a user based on connection of direct links through at least one core concept not selected by the user; and

communicating identified links to the user.

28. The method of claim 27, wherein said finding direct links includes detecting co-occurrence by employing a mutual information technique including the Fisher exact test to obtain a statistical P value expressing a significance of a detected co-occurrence.

29. The method of claim 27, wherein said finding direct links includes:

identifying an alias of a core concept in document contents; and

equating occurrence of the alias with occurrence of the core concept.

30. The method of claim 27, further comprising maintaining pointers between direct links and documents in which the direct links are found.

31. The method of claim 27, further comprising providing pointers to documents to the user, wherein the documents relate to an identified link.

32. The method of claim 27, wherein said finding direct links includes employing multiple, discipline-focused lexica organized according to the core concepts and identifying aliases by which the core concepts may be found in document contents.

33. The method of claim 27, further comprising:

communicating selectable lexica to the user;

receiving lexicon selections and initial search terms from the user;

extracting aliases from the initial search terms;

identifying candidate core concepts in lexica selected by the user based on the extracted aliases; and

communicating the candidate core concepts to the user for final selection.

34. The method of claim 27, further comprising receiving core concept selections and a specified depth of link from a user.

35. The method of claim 27, further comprising automatically identifying a type of a link between two core concepts based on automatic detection in link-related document contents of one of plural, predefined, candidate relationships between predefined categories associated with the two core concepts.

36. The method of claim 27, further comprising automatically identifying a direction of a link between two core concepts based on a type of the link between the two core concepts and predefined categories associated with the two core concepts.

37. The method of claim 27, further comprising rendering a graphic display of links between core concepts, with nodes corresponding to core concepts and edges corresponding to links.

38. The method of claim 27, further comprising rendering a graphic display of links between core concepts, wherein nodes serve as hyper links to summaries of information relating to associated core concepts, and edges serve as hyperlinks to collections of pointers to documents supporting associated links.

39. The method of claim 27, further comprising rendering a graphic display of links between core concepts, wherein edges between bounding nodes representing core concepts have visual characteristics identifying at least one of a strength of relationship between bounding nodes, a type of relationship between bounding nodes, and a direction of relationship between bounding nodes.

40. The method of claim 27, further comprising:

selecting a constraint list of candidate relationship types based on predefined categories associated with two core concepts bounding a direct link; and

automatically identifying a type of relationship associated with the direct link by finding occurrences of constraint list elements in proximity to detected co-occurrences of the two core concepts in document contents supporting the direct link.

41. The method of claim 27, further comprising applying a predefined direction associated with a candidate relationship between two core concepts of different predefined categories to a direct link bounded by the two core concepts.

42. The method of claim 27, further comprising automatically identifying a direction associated with a direct link between two core concepts of an identical type by matching a semantic template associated with a candidate relationship between the two core concepts to document contents in proximity to detected co-occurrences of the two core concepts in document contents supporting the direct link.

43. The method of claim 27, further comprising employing a gene lexicon organized according to core concepts corresponding to at least one of gene functions, protein functions, gene names, protein names, gene structures, and protein structures.

44. The method of claim 27, further comprising employing multiple, discipline-focused lexica organized according to core concepts pertaining to respective research disciples, including employing a gene lexicon, a disease lexicon, a drug lexicon, a tissue lexicon, and a taxonomy lexicon.